Dynamic Delirium Prediction in the Intensive Care Unit using Machine Learning on Electronic Health Records

Miguel Contreras[1,3], Brandon Silva[1,3], Benjamin Shickel[2,3], Sabyasachi Bandyopadhyay[1,3], Ziyuan Guan[2,3], Yuanfang Ren[2,3], Tezcan Ozrazgat-Baslanti[2,3], Kia Khezeli[1,3], Azra Bihorac[2,3], and Parisa Rashidi[1,3]

[1]Department of Biomedical Engineering, University of Florida, Gainesville, FL USA
[2]Department of Medicine, University of Florida, Gainesville, FL USA
[3]Intelligent Critical Care Center (IC3), University of Florida, Gainesville, FL USA

Abstract— **Delirium is a syndrome of acute brain failure which is prevalent amongst older adults in the Intensive Care Unit (ICU). Incidence of delirium can significantly worsen prognosis and increase mortality, therefore necessitating its rapid and continual assessment in the ICU. Currently, the common approach for delirium assessment is manual and sporadic. Hence, there exists a critical need for a robust and automated system for predicting delirium in the ICU. In this work, we develop a machine learning (ML) system for real-time prediction of delirium using Electronic Health Record (EHR) data. Unlike prior approaches which provide one delirium prediction label per entire ICU stay, our approach provides predictions every 12 hours. We use the latest 12 hours of ICU data, along with patient demographic and medical history data, to predict delirium risk in the next 12-hour window. This enables delirium risk prediction as soon as 12 hours after ICU admission. We train and test four ML classification algorithms on longitudinal EHR data pertaining to 16,327 ICU stays of 13,395 patients covering a total of 56,297 12-hour windows in the ICU to predict the dynamic incidence of delirium. The best performing algorithm was Categorical Boosting which achieved an area under receiver operating characteristic curve (AUROC) of 0.87 (95% Confidence Interval; C.I, 0.86-0.87). The deployment of this ML system in ICUs can enable early identification of delirium, thereby reducing its deleterious impact on long-term adverse outcomes, such as ICU cost, length of stay and mortality.**

Clinical Relevance— **This ML system can dynamically predict occurrence of delirium within the next 12 hours using the latest 12 hours of ICU data, along with patient demographic and medical history data.**

I. Introduction

Delirium is defined as a syndrome of acute brain failure characterized by deficits in attention, awareness, and cognition that fluctuate in severity over time [1]. This condition has a high incidence in patients in the intensive care unit (ICU), with an incidence of over 80% having been reported for mechanically-ventilated patients [2], [3]. It is associated with higher mortality risk [4], longer hospital stays [3], and long-term cognitive impairment [5]. Current approaches for delirium diagnosis are limited to methods which assess brain status or conduct an overall risk assessment but do not allow for early or continual dynamic diagnosis of delirium in the ICU. Such approaches include the use of scores such as the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) [6], Intensive Care Delirium Screening Checklist (ICDSC) [7], neuroimaging such as Magnetic Resonance Imaging (MRI), and biomarkers such as S100 calcium-binding protein B (S100B) and Neuron-Specific Enolase (NSE) [8]. These approaches are limited in their applicability as manual delirium evaluation has low implementation rates and the assessments are often not individualized for each patient [9]. Since a patient's condition is rapidly evolving in the ICU, it is imperative to capture this dynamic behavior to develop patient-centric, granular delirium assessments for planning effective interventions.

Given the wealth of data available in modern Electronic Health Record (EHR) systems, including vital signs, medications, laboratory results, assessment scores, previous diagnoses, and demographic information, there is an opportunity to use data-driven approaches to dynamically predict delirium in the ICU. Machine Learning (ML) has previously been used to develop multiple dynamic risk assessment scores for different time-sensitive outcomes in the ICU such as sepsis, organ failure, antibiotic administration and mortality [10]–[13].

Multiple ML approaches have also been applied to delirium prediction in the ICU. In particular, some studies have predicted the overall risk of delirium during ICU stay using clinical variables and risk factors available within the first 24 hours of admission [14], [15]. Other studies have focused on predicting delirium within the first 24 hours of ICU stay using clinical variables available within 4 hours after ICU admission [16]. Additionally, some studies have performed postoperative delirium prediction using preoperative data [17]. Although these studies have shown the feasibility of applying ML to delirium prediction, they do not consider the problem of "dynamic delirium prediction". A patient could shift from a non-delirious state to a delirious state and vice versa depending on the accumulation or addition of risk factors at different times during their ICU stay [18]. A dynamic delirium prediction system would enable caregivers to adjust their interventions in a timely manner.

Research supported by NIH/NINDS R01 NS120924, NIH/NIBIB R01 EB029699, NIH/NIGMS R01 GM110240, and NIH Bridge2AI OT2OD032701, as well as the NIH/NCATS Clinical and Translational Sciences Award to the University of Florida UL1 TR000064. P.R. was supported by NSF CAREER 1750192

Corresponding author: Parisa Rashidi (parisa.rashidi@ufl.edu)

In this study, we propose an ML system for dynamic prediction of delirium in the ICU. Closer to our setting, Lucini et al. [19] developed a dynamic delirium prediction for 12-hour and 24-hour windows but the model requires at least 24 hours of ICU data. Prediction of delirium within the first 24 hours of ICU admission is of great importance as many patients can present it within this time window [20], [21]. This study makes use of the latest 12 hours of ICU data along with patient's information at admission and medical history, making it possible to predict delirium as soon as after 12 hours of ICU admission and continues to make predictions every 12 hours to dynamically evaluate the evolving risk of delirium for a patient.

Our contribution is the development of the first delirium real-time predictive system based on 12-hour observations, which achieves comparable performance to existing methods. Four ML algorithms were tested for dynamic delirium prediction: Random Forest (RF), Extreme Gradient Boosting (XGB), Categorical Boosting (CB), and Gated Recurrent Unit (GRU). Other recent deep learning algorithms were tested as well but achieved similar performance. Furthermore, the tested ML algorithms provide the advantage of not requiring high computational resources and being easy to deploy in real-time. We hypothesize that capturing changes in clinical variables in the latest 12 hours, along with risk factors present at the time of admission, will accomplish accurate delirium risk prediction in the next 12-hour window.

II. Methodology

A. Data Collection

The data used in this study was collected from patients in the ICU at the University of Florida (UF) Health Shands Hospital (IRB number: IRB201901123), from 2012 to 2019, comprising a total of 54,467 patients, 72,902 hospital admissions, and 77,502 ICU stays. In this study, an ICU stay is defined as a unique admission to an ICU unit for a patient, meaning one patient can have multiple ICU stays. If there were multiple ICU stays for one patient within a 24-hour period, they were merged into a single ICU stay. For inclusion in the study, ICU stays needed to have a duration of at least 12 hours and at least one CAM-ICU score recorded during the stay. For each ICU stay, the delirium status was determined once every 12 hours through CAM-ICU scoring. A patient was considered delirious during a 12-hour window if they had at least one positive CAM-ICU score during that window. Additionally, ICU stays which were missing recordings for all clinical variables for more than 1 hour were removed. This resulted in a final cohort of 13,395 patients with 15,633 hospital admissions and 16,327 ICU stays. The cohort consisted of a total of 56,297 12-hour shifts from all ICU stays, of which 12,871 presented a delirium episode (23% incidence). The dataset was split into development and evaluation sets based on patient IDs, to avoid patient overlap between sets, using an 80:20 ratio. The development set consisted of 10,716 patients with 12,506 hospital admissions and 13,061 ICU stays. The evaluation set consisted of 2,679 patients with 3,127 hospital admissions and 3,266 ICU stays. The cohort selection process is summarized in Figure 1.

B. Feature extraction

The features used for development of the ML algorithms included both temporal and static data. The temporal data was extracted from the EHR system and consisted of 6 vital signs measurements (heart rate, systolic blood pressure, diastolic blood pressure, oxygen flow rate, respiratory rate), a total of 153 unique medications and 92 laboratory results, which were obtained after removing medications and lab tests present in less than 5% of ICU stays, and 30 ICU assessment scores and sub-scores (*e.g.*, both Braden score and its six component sub-scores were included). This resulted in a total of 281 temporal features. The static data consisted of 501 patient characteristics including demographics, admission information, comorbidities, admission type, neighborhood characteristics, and patient history from the previous year which included medications and laboratory tests. The final number of features combining temporal and static data was 782.

C. Data Processing

The temporal data was tabularized such that each column represented one temporal feature. Missing values were imputed by first performing linear interpolation on individual 12-hour windows and propagating backward and forward values where the conditions for interpolation were not satisfied. Remaining missing values were then imputed using the average for each feature based on the development set. For medication variables, zero was used for imputation of missing values since medications are either given or not given at specific timepoints. All values were then scaled using MinMax scaling [22].

D. Machine Learning

Four ML algorithms were investigated: Random Forest (RF), Extreme Gradient Boosting (XGB), Categorical Boosting (CB), and Gated Recurrent Unit (GRU). For RF, CB, and XGB algorithms, statistical features were calculated to convert the temporal features into a static space. These features included: mean, median, minimum, maximum, standard deviation, first value, last value, 25^{th} percentile, 75^{th} percentile, interquartile range, minimum to maximum range, minimum to last value difference, maximum to last value difference, average difference between timesteps, skewness, and kurtosis. Calculation of these statistical features resulted in a total of 4,496 features (281 temporal features × 16 statistical features), which when added with static features resulted in 4,997 final features. For the GRU algorithm, the temporal data was fed through a single GRU layer to obtain a temporal embedding, while the static data was fed through a fully connected layer to obtain a static embedding. Both embeddings were then concatenated and fed through two fully connected layers and a final output layer. The output for all algorithms was a single score representing the risk of delirium for the next 12 hours. The processing workflow is summarized in Figure 2.

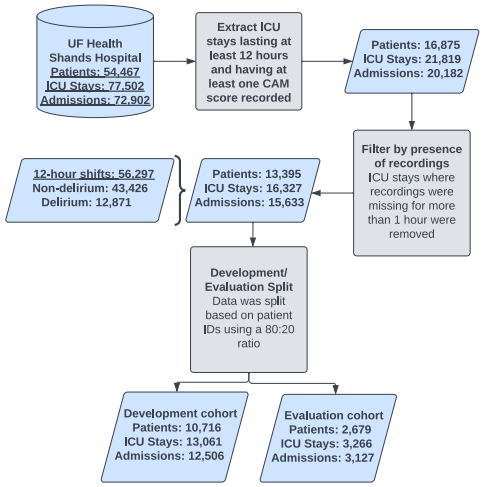

Figure 1. Diagram representing workflow for selection of cohort for this study.

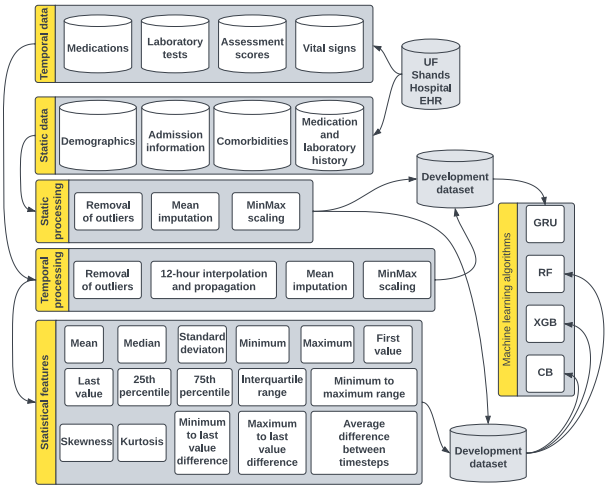

Figure 2. Diagram representing workflow for data processing pipeline for ML algorithm input.

E. Model Training and Evaluation

The ML algorithms were trained using 5-fold cross validation on the development set where hyperparameters were selected using the maximum sum of fold-wise area under the receiving operator characteristic curve (AUROC) and area under the precision-recall curve (AUPRC). The evaluation set was then used to assess final performance of the models by performing a 100-iteration bootstrap to calculate the 95% confidence interval (CI) for each algorithm. Besides AUROC and AUPRC, four other metrics were used for evaluation: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). These metrics were calculated by finding the optimal classification threshold using Youden Index on the receiving operator curve (ROC) [23]. For determining if the difference in performance was statistically significant, AUROC and AUPRC values between algorithms were compared using a two-sample t-test. Furthermore, SHapley Additive exPlanations (SHAP) [24] analysis was performed to determine the most relevant features that the best model used for making predictions.

III. Results

The results for all ML algorithms are summarized in Table I. CB demonstrated higher ability than all other algorithms ($p << 0.01$ for AUROC and AUPRC compared to all other algorithms), with AUROC 0.87 (0.86-0.87), AUPRC 0.62 (0.59-0.64), sensitivity 0.88 (0.86-0.91), specificity 0.72 (0.69-0.75), PPV 0.49 (0.46-0.51), and NPV 0.96 (0.95-0.96). The ROC and precision-recall curve (PRC) for all algorithms were plotted and compared to a no-skill model (Figure 3).

Results from SHAP analysis on CB algorithm are shown in Figure 4. The 15 most important features as determined by this analysis included some ICU assessment scores (*e.g.,* Morse, Glasgow Coma Scale (GCS), and Braden scores) as well as patient and admission information (*e.g.*, age, hospital service department where patient was admitted, attending doctor, and referring doctor).

IV. Discussion

The results presented in this study show the development of a dynamic delirium prediction system in the ICU. This system makes use of both temporal variables measured in the ICU and static variables obtained at admission, making use of ML to predict delirium risk in the next 12 hours. The best ML model, CB, showed comparable performance to state-of-the-art methods for delirium prediction in recent literature, achieving 0.87 AUROC compared to 0.91 AUROC in [19]. The main difference between these studies is the number of patients (38,436 patients compared to 13,395 patients in our cohort). However, the present study provides the advantage of requiring only the latest 12 hours of ICU data along patient and admission information, whereas [19] required at least 24 hours of data for an ICU stay to make a prediction. Predicting delirium risk within the first 24 hours of ICU admission is relevant. In the initially extracted cohort (Figure 1), among 4,000 patients who had delirium during their ICU stay, 29% (1,147 patients) and 40% (1,606 patients) developed delirium in the first 12 and 24 hours of their admission to the ICU, respectively. Furthermore, in the final cohort used for this study, among 3,324 patients who had delirium during their ICU stay, 22% (737 patients) presented delirium sometime between 12-24 hours after ICU admission.

Compared to other studies using ML for delirium prediction, our results show improvement in terms of AUROC with regards to studies using similar or even larger cohorts [14], [17]. Furthermore, the top features recognized through SHAP analysis were consistent with features found in other studies [14], [17]. The influence of age, attending doctor, and hospital admitting service have been shown to have an impact on delirium risk [14]. Braden scores assess sensory perception, mobility and activity of the patient which are all correlated to potential development of delirium [25]. Glasgow Coma Scale (GCS), which is used to assess presence of coma, has also been shown to be correlated to delirium given it evaluates the mental state of the patient [26]. Morse fall scale, which contains subcomponents such as mental status, assesses fall

TABLE I: RESULTS FOR DELIRIUM PREDICTION ON EVALUATION SET

	RF	*XGB*	*CB*	*GRU*
AUROC (95% CI)	0.85 (0.84-0.86)	0.85 (0.85-0.86)	**0.87 (0.86-0.87)**	0.84 (0.83-0.84)
AUPRC (95% CI)	0.59 (0.57-0.61)	0.59 (0.57-0.61)	**0.62 (0.59-0.64)**	0.53 (0.51-0.54)
Sensitivity (95% CI)	0.85 (0.82-0.88)	0.87 (0.84-0.90)	**0.88 (0.86-0.91)**	0.86 (0.83-0.89)
Specificity (95% CI)	0.72 (0.70-0.74)	0.71 (0.68-0.74)	**0.72 (0.69-0.75)**	0.69 (0.66-0.72)
PPV (95% CI)	0.47 (0.45-0.49)	0.47 (0.45-0.50)	**0.49 (0.46-0.51)**	0.45 (0.43-0.47)
NPV (95% CI)	0.94 (0.93-0.95)	0.95 (0.94-0.96)	**0.96 (0.95-0.96)**	0.94 (0.93-0.96)

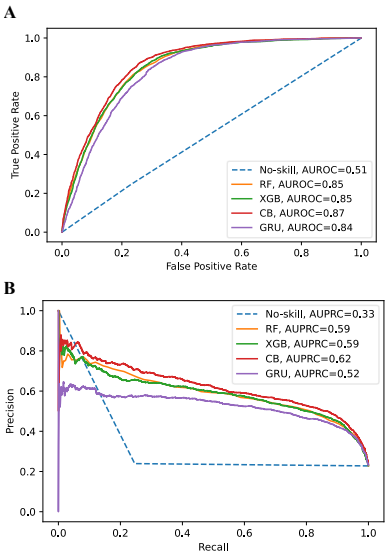

Figure 3. (A) ROC curves for all algorithms where x-axis represents the false positive rate and y-axis represents the true positive rate. (B) PRC curves for all algorithms where x-axis represents the recall (*i.e.,* sensitivity) and y-axis represents the precision (*i.e.,* PPV).

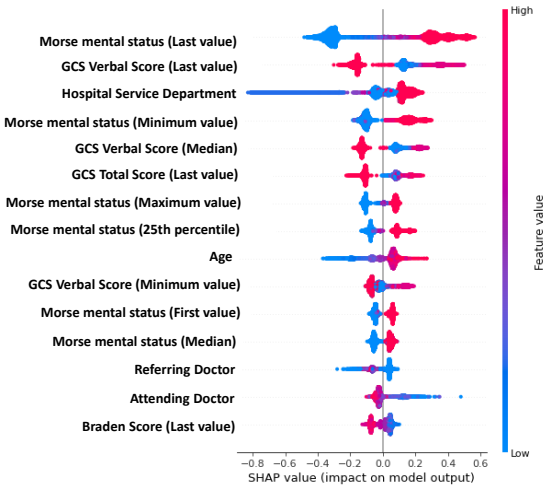

Figure 4. SHAP analysis on CB algorithm showing top 15 features, ordered by descending importance (top to bottom). Horizontal axis represents SHAP value determining the impact of a feature on model output. Colder colors represent lower feature values while warmer colors represent higher feature values.

risk for a patient, which has been correlated to delirium [27], all of which is consistent with our findings.

The usage of AUPRC in this study provides an additional metric to account for class imbalance. As seen in Figure 2B, CB achieved the highest AUPRC (0.62) amongst all ML algorithms which was significantly higher ($p << 0.01$) when compared to a no-skill model (0.33 AUPRC). Furthermore, CB achieved the highest specificity (0.72) and sensitivity (0.88) compared to all other ML algorithms. A higher sensitivity is important for reducing the risk of an inflated false-positive rate due to repeated measurement of delirium status of the same patient. The low value for PPV could be explained by the degree to which each patient experiences delirium, from mild to severe. The CAM-ICU assessment method is known for being more specific than sensitive showing a pooled 80% sensitivity and 95% specificity in a meta-analysis of 9 studies [28]. Furthermore, CAM-ICU has also been shown to have low sensitivity for patients with mild delirium [29]. Therefore, the high AUROC and low PPV could indicate presence of patients in the cohort that had mild delirium but were not captured by the CAM-ICU assessment.

The main limitations of this study include lack of external validation on data from other health centers and the non-usage of state-of-the-art ML algorithms for time series classification. Since the primary aim of this study was to develop an easy-to-deploy dynamic delirium prediction system using EHR data over 12-hr windows, we used well-established ML algorithms and processing pipelines to achieve this objective. Future work will include external validation of this model on publicly available clinical data (such as MIMIC-IV [30] and eICU [31]), and use Transformer models for time-series prediction.

V. CONCLUSION

Early detection of delirium in the ICU is critical for reducing delirium-related adverse events. Manual assessment is slow in determining delirium and can only catch delirium after the onset of symptoms. In this work, we established the efficacy of ML algorithms to predict delirium episodes 12 hours in advance using EHR data from ICU patients at the UF Health Shands Hospital. Real-time deployment of these algorithms has the potential to provide early warnings to the clinical staff of delirium onset, which can help reduce healthcare costs and save lives. Future work will build upon our methods and validate our approaches using publicly available datasets to ensure the robustness of the algorithms and their generalizability to other hospital settings.

References

[1] R. P. Thom, N. C. Levy-Carrick, M. Bui, and D. Silbersweig, "Delirium," *Am. J. Psychiatry*, vol. 176, no. 10, pp. 785–793, Oct. 2019.

[2] E. W. Ely *et al.*, "Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU)," *JAMA*, vol. 286, no. 21, pp. 2703–2710, Dec. 2001.

[3] E. Ely *et al.*, "The impact of delirium in the intensive care unit on hospital length of stay," *Intensive Care Med.*, vol. 27, no. 12, pp. 1892–1900, Dec. 2001.

[4] E. Pauley, A. Lishmanov, S. Schumann, G. J. Gala, S. van Diepen, and J. N. Katz, "Delirium is a robust predictor of morbidity and mortality among critically ill patients treated in the cardiac intensive care unit," *Am. Heart J.*, vol. 170, no. 1, pp. 79-86.e1, Jul. 2015.

[5] T. E. Goldberg *et al.*, "Association of Delirium With Long-term Cognitive Decline: A Meta-analysis," *JAMA Neurol.*, vol. 77, no. 11, pp. 1373–1381, Nov. 2020.

[6] E. W. Ely *et al.*, "Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU)," *Crit. Care Med.*, vol. 29, no. 7, pp. 1370–1379, Jul. 2001.

[7] N. E. Brummel, E. E. Vasilevskis, J. H. Han, L. Boehm, B. T. Pun, and E. W. Ely, "Implementing delirium screening in the ICU: secrets to success," *Crit. Care Med.*, vol. 41, no. 9, pp. 2196–2208, Sep. 2013.

[8] F. Esen, G. Orhun, P. E. Özcan, A. R. Brenes Bastos, and E. Tüzün, "Diagnosing acute brain dysfunction due to sepsis," *Neurol. Sci.*, vol. 41, no. 1, pp. 25–33, Jan. 2020.

[9] K. Ishii *et al.*, "Current status of delirium assessment tools in the intensive care unit: a prospective multicenter observational survey," *Sci. Rep.*, vol. 12, no. 1, Art. no. 1, Feb. 2022.

[10] B. Shickel, T. J. Loftus, L. Adhikari, T. Ozrazgat-Baslanti, A. Bihorac, and P. Rashidi, "DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning," *Sci. Rep.*, vol. 9, p. 1879, Feb. 2019.

[11] B. Shickel *et al.*, "Multi-dimensional patient acuity estimation with longitudinal EHR tokenization and flexible transformer networks," *Front. Digit. Health*, vol. 4, 2022.

[12] D. A. Kaji *et al.*, "An attention based deep learning model of clinical events in the intensive care unit," *PloS One*, vol. 14, no. 2, p. e0211057, 2019.

[13] S. Y. Kim *et al.*, "A deep learning model for real-time mortality prediction in critically ill children," *Crit. Care*, vol. 23, no. 1, p. 279, Aug. 2019.

[14] A. Davoudi, A. Ebadi, P. Rashidi, T. Ozrazgat-Baslanti, A. Bihorac, and A. C. Bursian, "Delirium Prediction using Machine Learning Models on Preoperative Electronic Health Records Data," *Proc. IEEE Int. Symp. Bioinforma. Bioeng.*, vol. 2017, pp. 568–573, Oct. 2017.

[15] A. Wong, A. T. Young, A. S. Liang, R. Gonzales, V. C. Douglas, and D. Hadley, "Development and Validation of an Electronic Health Record–Based Machine Learning Model to Estimate Delirium Risk in Newly Hospitalized Patients Without Known Cognitive Impairment," *JAMA Netw. Open*, vol. 1, no. 4, p. e181018, Aug. 2018.

[16] S. Hur, R.-E. Ko, J. Yoo, J. Ha, W. C. Cha, and C. R. Chung, "A Machine Learning–Based Algorithm for the Prediction of Intensive Care Unit Delirium (PRIDE): Retrospective Study," *JMIR Med. Inform.*, vol. 9, no. 7, p. e23401, Jul. 2021.

[17] A. Bishara *et al.*, "Postoperative delirium prediction using machine learning models and preoperative electronic health record data," *BMC Anesthesiol.*, vol. 22, no. 1, p. 8, Jan. 2022.

[18] H. Fan *et al.*, "Development and validation of a dynamic delirium prediction rule in patients admitted to the Intensive Care Units (DYNAMIC-ICU): A prospective cohort study," *Int. J. Nurs. Stud.*, vol. 93, pp. 64–73, May 2019.

[19] F. R. Lucini, H. T. Stelfox, and J. Lee, "Deep Learning–Based Recurrent Delirium Prediction in Critically Ill Patients," *Crit. Care Med.*, vol. Publish Ahead of Print, Feb. 2023.

[20] E. W. Ely *et al.*, "Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit," *JAMA*, vol. 291, no. 14, pp. 1753–1762, Apr. 2004.

[21] R. B. Serafim *et al.*, "Delirium in postoperative nonventilated intensive care patients: risk factors and outcomes," *Ann. Intensive Care*, vol. 2, no. 1, p. 51, Dec. 2012.

[22] "sklearn.preprocessing.MinMaxScaler," *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html.

[23] R. Fluss, D. Faraggi, and B. Reiser, "Estimation of the Youden Index and its associated cutoff point," *Biom. J. Biom. Z.*, vol. 47, no. 4, pp. 458–472, Aug. 2005.

[24] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017.

[25] A. Khan, K. Heslin, M. Simpson, and M. L. Malone, "Can Variables From the Electronic Health Record Identify Delirium at Bedside?," *J. Patient-Centered Res. Rev.*, vol. 9, no. 3, pp. 174–180, Jul. 2022.

[26] B. Van Rompaey, M. M. Elseviers, M. J. Schuurmans, L. M. Shortridge-Baggett, S. Truijen, and L. Bossaert, "Risk factors for delirium in intensive care patients: a prospective cohort study," *Crit. Care*, vol. 13, no. 3, p. R77, May 2009.

[27] B. Kalivas *et al.*, "The Association between Delirium and In-Hospital Falls: A Cross-Sectional Analysis of a Delirium Screening Program," *J. Aging Res.*, vol. 2023, p. e1562773, Jan. 2023.

[28] D. Gusmao-Flores, J. I. F. Salluh, R. Á. Chalhub, and L. C. Quarantini, "The confusion assessment method for the intensive care unit (CAM-ICU) and intensive care delirium screening checklist (ICDSC) for the diagnosis of delirium: a systematic review and meta-analysis of clinical studies," *Crit. Care Lond. Engl.*, vol. 16, no. 4, p. R115, Jul. 2012.

[29] S. Boettger *et al.*, "Delirium in the intensive care setting: A reevaluation of the validity of the CAM–ICU and ICDSC versus the DSM–IV–TR in determining a diagnosis of delirium as part of the daily clinical routine," *Palliat. Support. Care*, vol. 15, no. 6, pp. 675–683, Dec. 2017.

[30] A. E. W. Johnson *et al.*, "MIMIC-IV, a freely accessible electronic health record dataset," *Sci. Data*, vol. 10, no. 1, Art. no. 1, Jan. 2023.

[31] T. J. Pollard, A. E. W. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, "The eICU Collaborative Research Database, a freely available multi-center database for critical care research," *Sci. Data*, vol. 5, no. 1, Art. no. 1, Sep. 2018.